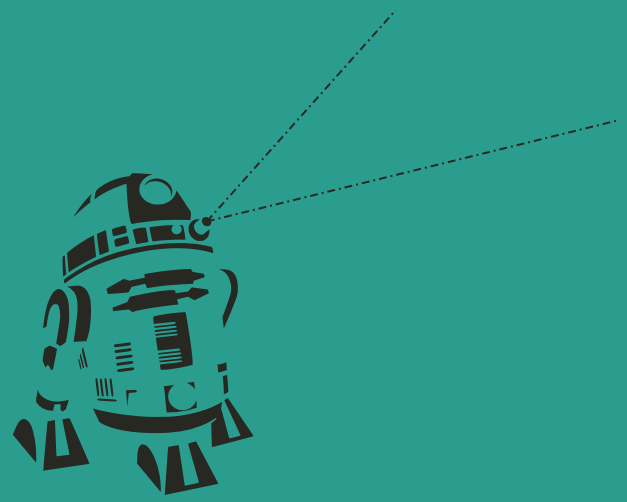
Introduction to Machine Learning

Albert Ruiz

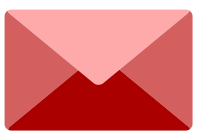
What was the first Machine Learning application?



First ML application: the spam filter

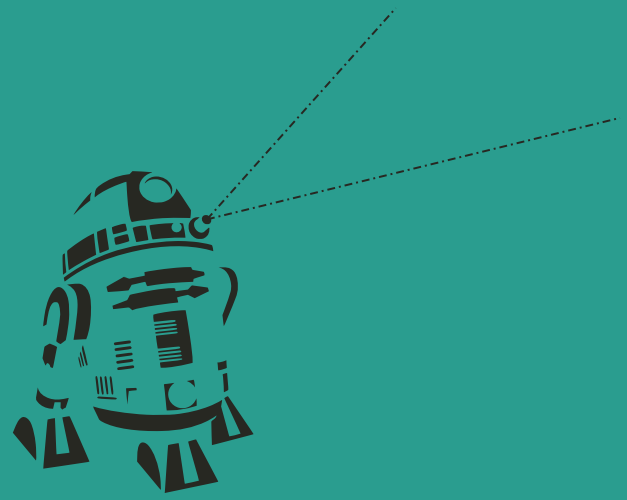
Ham or Spam?

✔



✘

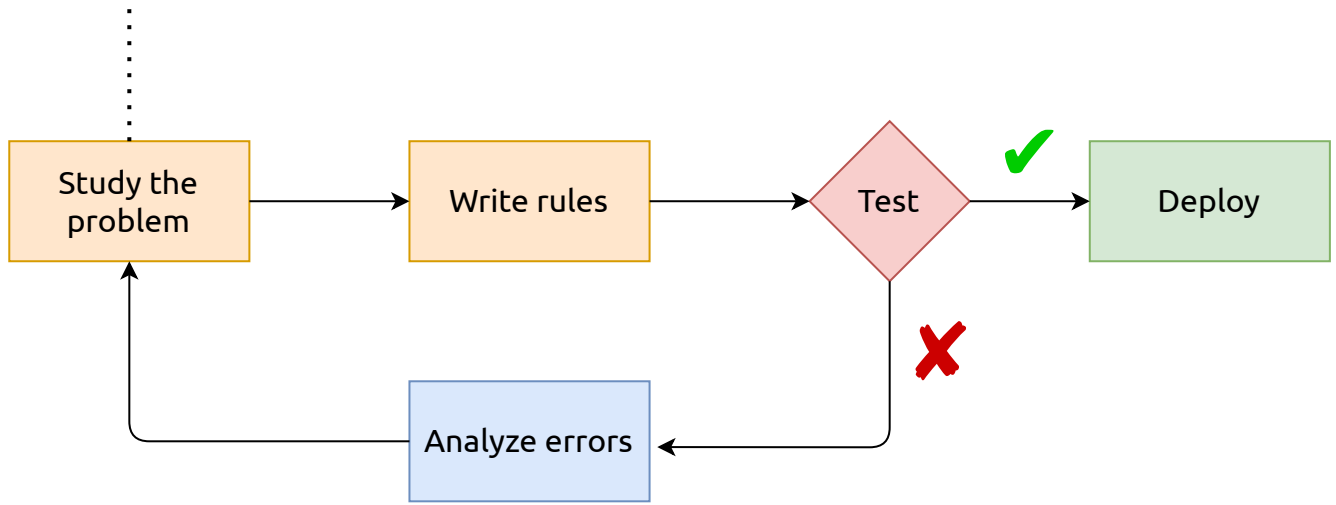
How would you code a spam filter?



Coding a spam filter: traditional approach

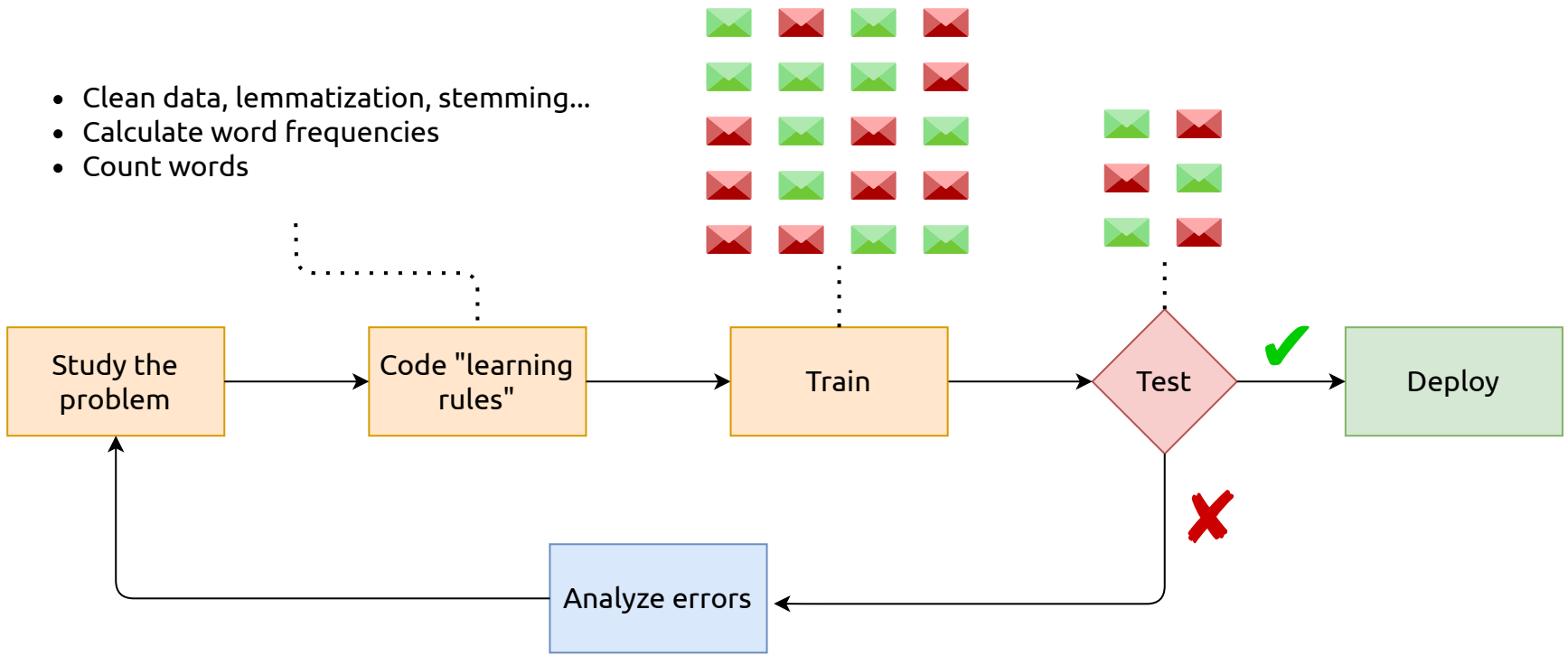
Developer does (example):

- Find common words: IBAN , discount , offer , bank , password ...
- Find patterns in introductory phrases: Dear Sir/Madam , Mr/Mrs/Miss ...
- Find patterns in email addresses: @hacker.com , @no-reply.com ...
- Calculate weights



Coding a spam filter: ML approach

- Clean data, lemmatization, stemming...
- Calculate word frequencies
- Count words



What are we going to learn today?

Agenda

1. Introduction (5 min)
    - What is ML?
    - Why ML?
    - ML in insurance
2. End-to-end ML (45 min)
    - Data
    - Processing
    - Vectorization
    - Modelling
    - Visualization
3. Hands-on ML (practice, 1h)

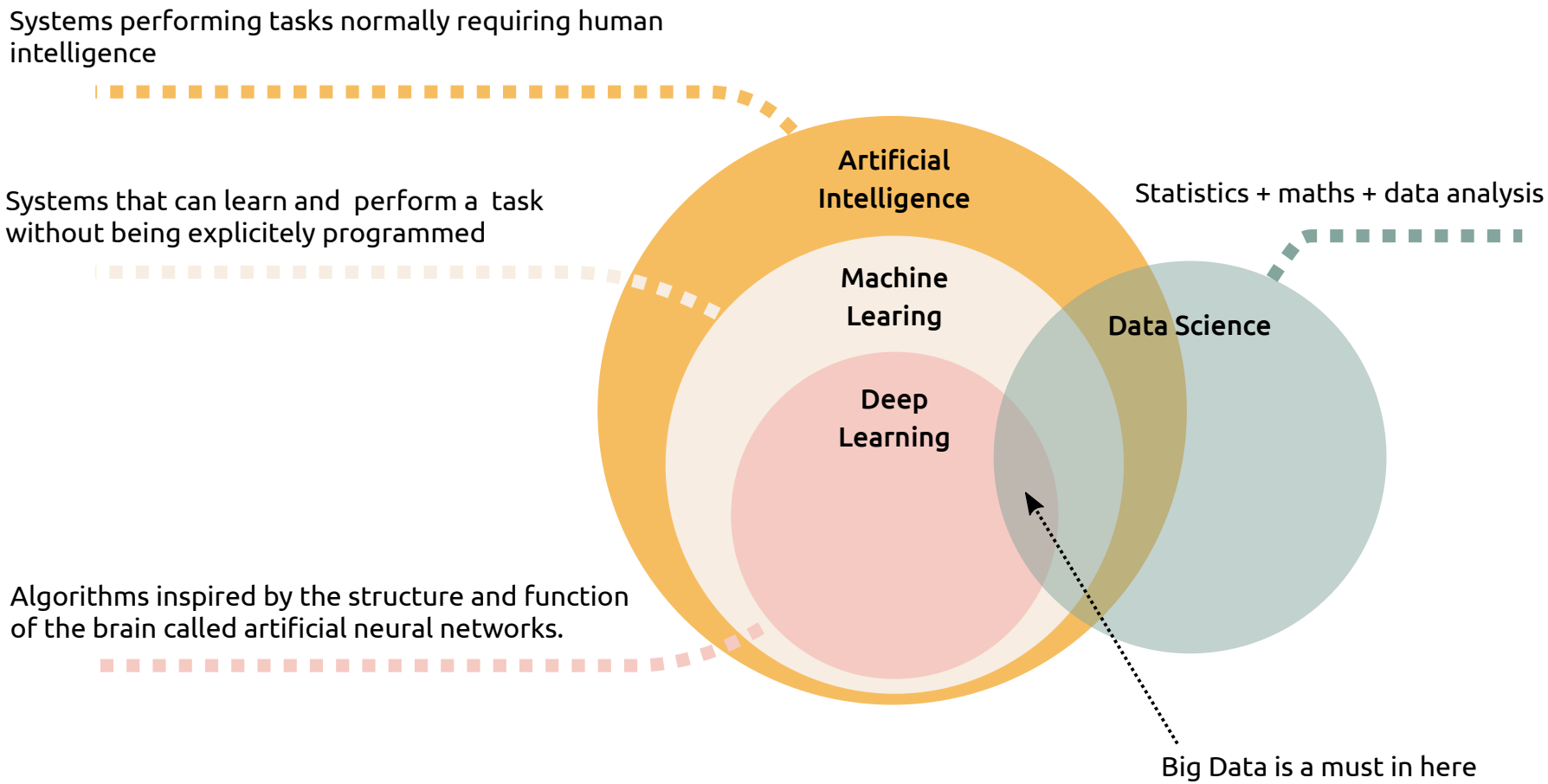
Introduction

What is ML?

Machine learning is the field of study that gives computers the ability to learn without being explicitely programmed.

[Arthur L. Samuel, 1959]

AI / ML / DL / DS / BD


Systems performing tasks normally requiring human intelligence
Systems that can learn and perform a task without being explicitely programmed
Algorithms inspired by the structure and function of the brain called artificial neural networks.
Artificial Intelligence
Machine Learing
Deep Learning
Data Science
Statistics + maths + data analysis
Big Data is a must in here

Why ML?

Some modern problems are too complex for traditional approaches:

- Problems that require fine-tuning or long list of rules
- Problems with fluctuating data
- Getting insights from large amounts of data

Machine Learning can help humans learn!

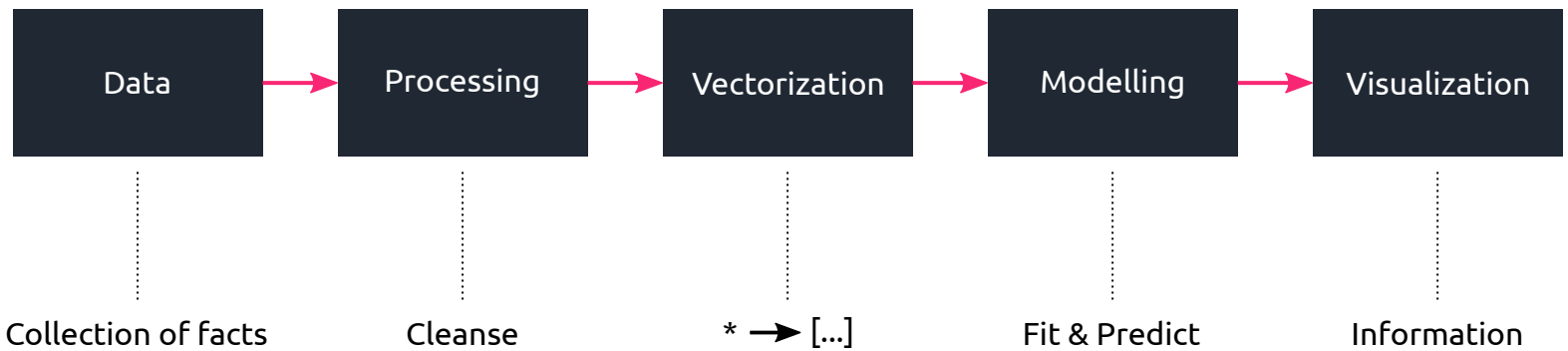
Examples of applications

- Text classification
- Sentiment analyisis
- Summarizing long text
- Data extraction from images
- Fraud detection
- Chatbots
- Client segmentation
- Recommending a product to a client
- Speech recognition
- Forecasting

ML applied to the insurance sector

Common steps in a ML project

The common steps



Structured / Unstructured

Structured data

- Well defined interface to access data
- Data is formatted
- SQL, XML, HTML, JSON, XLSX, CSV, PNG...

Unstructured data

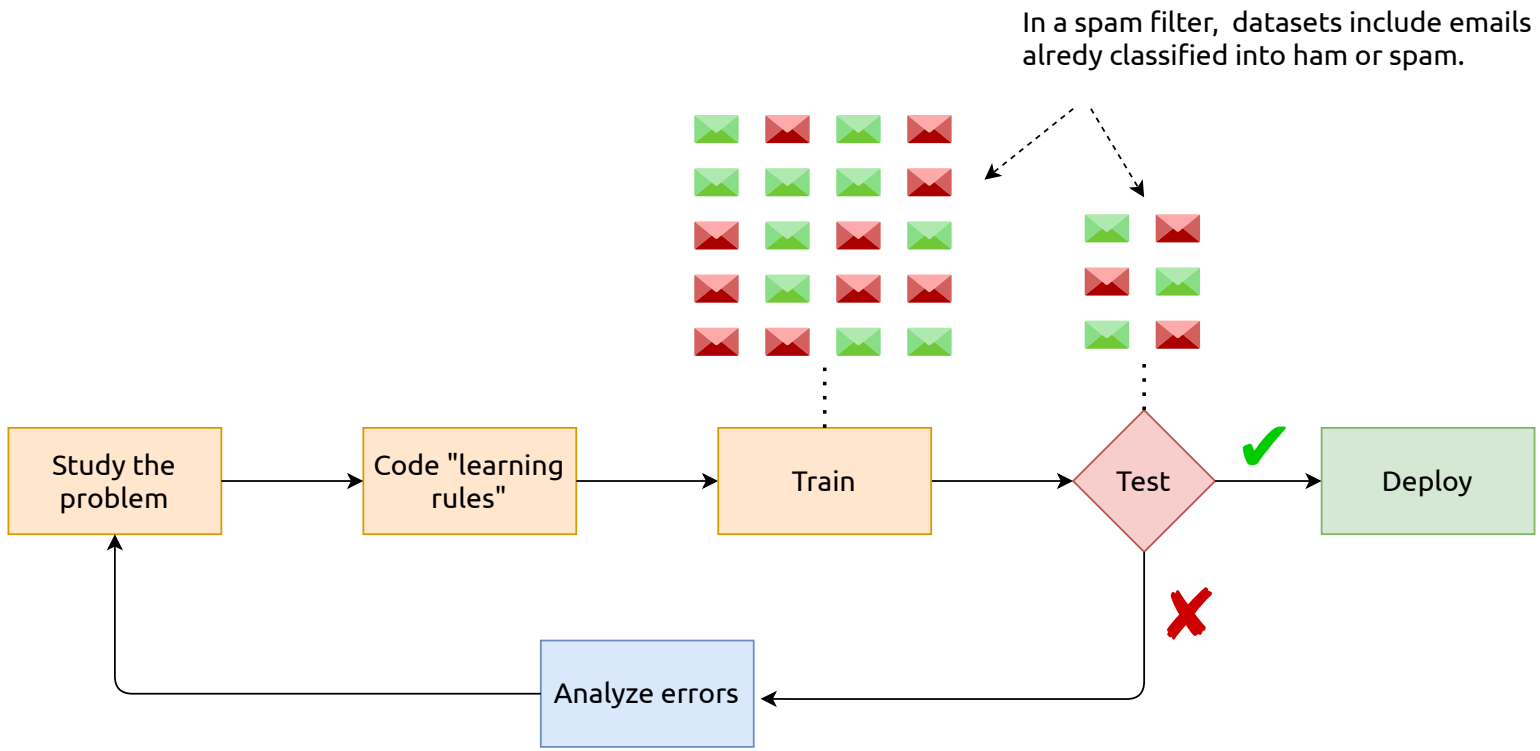
- No predefined interface to access data
- No predefined format
- PDF, TXT, TEX, MD...

When structured becomes unstructred

name	surname	sex	birthdate	birthplace	country	phone
Max	Rockatasnky	M	10-11-1984	Perth	AU	+61 8 6245 2100
Immortan	Joe	m	01-02-1949	Canberra	AU	+61 4 1234 5678
James	Connor	M	1985-02-28	Los Angeles	USA	unknown
Alex Murphy		M	1979	Detroit	US	tbc
John	McClane	M	1969-07-17	Los Angeles	US	4242706247
Pete	Mitchell	MALE	1972-10-10	San Diego	US	tbc

Labelled / Unlabelled

Labelled datasets include the desired solutions (labels), used in supervised learning.



Categorical / Quantitative

Training dataset / Testing dataset

Insuficient quantity of training data

Non-representative training data

Data

Processing

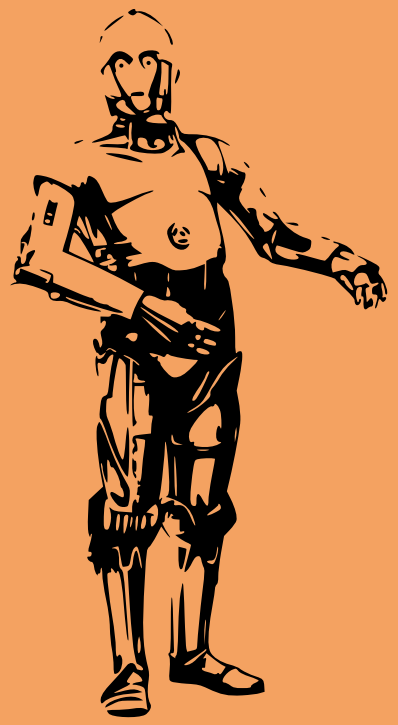
Vectorization

Modelling

Visualization

Poor quality

Hands-on ML (practice)



Questions?

(albert.ruizalvarez@zurich.com)

Thank you!